

Übung 17: CQP

Indexierung eines Korpus

1. Wechseln Sie auf den Rechner ella und kopieren Sie die Korpusdatei `/home/public/vorkurs_ss18/tiger.cwb` in Ihr Vorkurs-Verzeichnis. Die Datei enthält die Texte aus dem Tiger-Corpus (Nachrichtenartikel) im cwb-Format.²
2. Erstellen Sie einen neuen Ordner `mycorpus` für die Korpusdaten im Binärformat sowie einen Ordner `myregistry` für die Registerdatei.
3. Kodieren Sie Ihr Korpus. Indexieren Sie es anschließend.

Anfragen stellen

1. Starten Sie nun CQP und lassen Sie sich alle Korpora anzeigen.
2. Wählen Sie Ihr Korpus aus und führen Sie nun die folgenden Anfragen aus:³
 - a) Alle Vorkommen von dem Wort ‘erfolgreich’.
 - b) Die ersten 20 Vorkommen von ‘Bank’.
 - c) Alle Interjektionen.
 - d) Alle Nominalphrasen bestehend aus einem Artikel, einem Adjektiv und einem Nomen.
 - e) Lassen Sie sich nun auch die POS-tags im Ergebnisfenster anzeigen.
 - f) Alle Wörter mit dem Lemma ‘sein’, außer ‘bin’, die nicht in der Grundform (tag VVINP) vorkommen.
 - g) Ein Adverb (ADV) am Satzanfang.
3. Auswertungen von Treffern
 - a) Fragen Sie nach dem Wort ‘meine’ und lassen Sie sich die Häufigkeitsverteilung der POS-Tags ausgeben.
 - b) Speichern Sie das Ergebnis einer Anfrage in der Variable `res` und lassen Sie sich dann den Inhalt der Variablen anzeigen.
 - c) Setzen Sie die `DataDirectory` auf ihr aktuelles Verzeichnis.
 - d) Wählen Sie das Korpus erneut aus und leiten Sie die Ausgabe von `res` in eine Textdatei um. (Wenn Sie cqp verlassen, sollte die Datei in Ihrem Verzeichnis liegen.)

²<http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger.html>

³Beschreibung des STTS Tag Sets: http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/stts_guide.pdf